SMOOTHING APPROXIMATIONS TO NONSMOOTH OPTIMIZATION PROBLEMS

X.Q. YANG[1]

Abstract

We study certain types of composite nonsmooth minimization problems by introducing a general smooth approximation method. Under various conditions we derive bounds on error estimates of the functional values of original objective function at an approximate optimal solution and at the optimal solution. Finally, we obtain second-order necessary optimality conditions for the smooth approximation prob lems using a recently introduced generalized second-order directional derivative.

1. Introduction

The study of optimization techniques for solving nonsmooth optimization problems has been a growing interest in recent years. It has been shown that the smoothing approximation techniques are efficient methods for solving certain specially structured nonsmooth problems, see [2, 3, 4, 6, 10, 13, 14]. El-Attar et al. [6] introduced a smoothing method for a nonlinear l_1-minimization problem using a square root of the sum of square of nonsmooth absolute value function and a smoothing parameter. The applications of their method for a class of nonsmooth optimization problems in finite dimensional space and for an exact penalty problem in infinite dimensional space were studied by Ben-Tal et al. [2] and Yang [16] respectively. Ben-Tal and Teboulle [1] obtained a smoothing method for a general nonsmooth problem using recession functions of convex analysis. Bertsekas [3, 4] examined a two-parameter approximation for a nonsmooth minimization problem with kinks and derived first-order necessary conditions via the necessary conditions of the smooth approximation problem. Teo and Goh [13], Jennings and Teo [10] and Teo et al. [14] studied a class of nonsmooth optimization problems by smoothing the nondifferentiable cost functional and nondifferentiable functional inequality constraints with parameter approximations.

[1]Department of Mathematics, University of Western Australia, Nedlands, W.A. 6009, Australia.

In this paper, we introduce a smoothing approximation method to study a general composite nonsmooth optimization problem which, for instance, covers nonlinear l_1-minimization problems and exact penalty function problems. The smoothing approximation is developed using a two-parameter approximation. One parameter guarantees the desired accuracy. The choice of second parameter allows us to cover a larger class of nonsmooth problems than the ones discussed in [2, 6, 16]. We show how second-order necessary conditions can be derived for the smooth approximation problem.

The outline of the paper is as follows. In Section 2, we introduce a smoothing approximation method to study a general composite nonsmooth optimization problem. In Section 3, we provide error estimate of the functional values of the original objective function at an approximate optimal solution and at the optimal solution under various conditions. In Section 4, we then obtain general second-order necessary optimality condition for the smooth approximation problem using a recently introduced generalized second-order directional derivative.

Let us first recall a generalized second-order directional derivative and a generalized chain rule introduced recently in [16, 17] which will be used in Section 4. Let X be a Banach space with the canonical pair $\langle \cdot, \cdot \rangle$ between X and its continuous dual space X^* and O be an open set of X. A real-valued function $f : O \longrightarrow \mathbb{R}$ is called Gâteaux differentiable on O if for every $x \in O$, there exists $Df(x) \in X^*$ such that for any $u \in X$

$$\lim_{s \downarrow 0} \frac{f(x + su) - f(x) - s\langle Df(x), u \rangle}{s} = 0.$$

A real-valued function $f : O \longrightarrow \mathbb{R}$ is called $C^{1,1}$ if f is Gâteaux differentiable on O and the gradient function is locally Lipschitz on O. In this case, the gradient of f is denoted as ∇f. Let $x \in O$. The generalized second-order directional derivative $f^{\infty}(x; u, v)$ of f at x in the direction $(u, v) \in X \times X$ is defined by

$$f^{\infty}(x; u, v) = \sup_{z \in X} \limsup_{s \downarrow 0} \frac{\langle \nabla f(x + sz + su), v \rangle - \langle \nabla f(x + sz), v \rangle}{s},$$

and the corresponding generalized Hessian of f at $x \in O$ is defined for each $u \in X$ by

$$\partial^{\infty} f(x)(u) = \{x^* \in X^* : f^{\infty}(x; u, v) \geq \langle x^*, v \rangle, \forall v \in X\}.$$

The function f is called twice semi-regular at $x \in O$ if for every $u, v \in X$ the limit

$$f''(x; u, v) = \lim_{\substack{s \downarrow 0 \\ t \downarrow 0}} \frac{f(x + su + tv) - f(x + su) - f(x + tv) + f(x)}{st}$$

exists and is equal to $f^{\infty}(x; u, v)$.

It is easy to see that a second-order necessary condition of f at a local minimum $a \in O$ is

$$f^{\infty\infty}(a; u, u) \geq 0, \qquad \forall u \in X.$$

The following useful chain rule for composite $C^{1,1}$ functions was established in [16]. Let $F = f \circ g$, where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_n) : O \subset X \longrightarrow \mathbb{R}^n$, $f, g_1, \ldots, g_n$ are $C^{1,1}$ functions. Then for each $u, v \in X$, $x \in O$, we have

$$F^{\infty\infty}(x; u, v) \leq f^{\infty\infty}(g(x); \nabla g(x)u, \nabla g(x)v) + h_x(u, v),$$

where

$$h_x(u, v) = \max \left\{ \sum_{i=1}^{n} \frac{\partial f(g(x))}{\partial x_i} \langle x_i^*, v \rangle : x_i^* \in \partial^{\infty\infty} g_i(x)(u), i = 1, \ldots, n \right\}.$$

2. A generalized smoothing problem

In this section, we introduce a smoothing approximation problem to study a general nonsmooth composite model problem which covers, for example, exact penalty function problems and nonlinear l_1-minimization problems (see, for example, [2, 5, 6, 7, 9, 11, 16]). We then compare the smoothing approximation problem with other smoothing problems given recently in [1] and see that our approximation problem does not depend on a recession function that is used in [1]. We begin by giving some examples.

EXAMPLE 2.1. Let $h_i : X \longrightarrow \mathbb{R}^p$ $(i = 1, \ldots, m)$ be differentiable functions. The following nonsmooth problem

$$\min_{x \in X} \sum_{i=1}^{m} \|h_i(x)\|$$

can be approximated by the smooth problem

$$\min_{x \in X} \sum_{i=1}^{m} \sqrt{\|h_i(x)\|^2 + \epsilon^2},$$

where $\|y\|$ is a norm in $\mathbb{R}^p$. For the case $p = 1$ and $X = \mathbb{R}^n$, this is the nonlinear l_1-minimization problem considered in El-Attar, Vidyasagar and Dutta [6]. Ben-Tal et al. [2] studied the case when $h_i(x) = Q_i x - r_i$, $X = \mathbb{R}^n$ and the norm $\|y\|$ is chosen as $\|y\| = \sqrt{y^T D y}$, where D is an $p \times p$ diagonal positive definite matrix, Q_i is a $p \times n$ matrix and $r_i \in \mathbb{R}^p$.

EXAMPLE 2.2. The (nonsmooth) exact penalty problem for the nonlinear programming problem:

$$\min_{x \in A} g_0(x)$$

(where $A = \{x \in X : x \in C, g_i(x) \geq 0, i = 1, \ldots, m\}$, C is a subset of X and $g_i : X \longrightarrow \mathbb{R}$ $(i = 0, 1, \ldots, m)$ are differentiable functions) is (for example, see [16])

$$\min_{x \in X} \ g_0(x) + M \left(\sum_{i=1}^{m} \max\{g_i(x), 0\} + d_C(x) \right),$$

where $M > 0$ is the penalty constant and the distance function $d_C(x)$ is defined by

$$d_C(x) = \inf\{\|x - y\| : y \in C\}.$$

This problem can be approximated by the smooth problem (see Yang [16])

$$\min_{x \in X} \ g_0(x) + M \left(\sum_{i=1}^{m} \sqrt{[\max\{g_i(x), 0\}]^2 + \epsilon^2} + \sqrt{d_C^2(x) + \epsilon^2} \right).$$

We now consider the nonsmooth optimization problem of the form:

$$\min_{x \in X} g(x), \ \text{where} \ g(x) = f(x) + G(f_1(x), \ldots, f_m(x)), \tag{P}$$

where X is a Banach space, and $f, f_i : X \longrightarrow \mathbb{R}, i = 1, \ldots, m, G : \mathbb{R}^m \longrightarrow \mathbb{R}$. This type of nonsmooth problem has been studied in [1, 2, 3, 7] under various conditions on the functions involved. In this paper, we suppose that the nonsmoothness of the problem (P) is, in general, due to the terms $f_i(x)$. Hence we assume that the functions satisfy the following conditions:

(A1) $f : X \longrightarrow \mathbb{R}$ is differentiable;

(A2) $G : \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable;

(A3) $f_i : X \longrightarrow \mathbb{R}_+, i = 1, \ldots, m$ are not, in general, differentiable, but for some $\alpha > 1$, the functions $f_i^\alpha, i = 1, \ldots, m$, are differentiable, where $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$.

These assumptions are used to smooth the original problem (P). For instance, the function G may take the form $\sum_{i=1}^{m} \beta_i y_i$. Some examples of the functions $f_i's$ are $\max\{g_i(x), 0\}$, and $|g_i(x)|$, where the functions $g_i : X \longrightarrow \mathbb{R}, i = 1, \ldots, m$ are differentiable.

For (P), we introduce the following approximation problem:

$$\min_{x \in X} g_\epsilon(x), \quad \text{where} \ g_\epsilon(x) = f(x) + G((f_1(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}, \ldots, (f_m(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}), \tag{P(ϵ)}$$

where $\epsilon > 0$ is a fixed smoothing parameter and $\alpha > 1$ is fixed. This is a smooth problem. This smooth problem may be solved by conventional methods, see [4]. The

use of α shows that our model is more useful than the ones given in [2, 6, 16] (with $\alpha = 2$), for example, let $f_1(x) = x^{\frac{1}{3}}$, then $f_1(x)^2$ is not differentiable, but $f_1(x)^3$ (with $\alpha = 3$) is differentiable.

It is worth recalling that Ben-Tal and Teboulle [1] considered the following nonsmooth problem:

$$\inf_{x \in \mathbb{R}^n} g(x), \text{ where } g(x) = F(f_1(x), \ldots, f_m(x)),$$

where $g(x) = F(f_1(x), \ldots, f_m(x))$, $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are real functions (for $i = 1, \ldots, m$) and $F : \mathbb{R}^m \longrightarrow \mathbb{R}$ is the function of the form:

$$F(y) = \sup\{h(x + y) - h(x) : x \in \text{dom}(h)\},$$

where h is a proper convex function and $\text{dom}(h) = \{x \in \mathbb{R}^m : h(x) < +\infty\}$. In fact, F is called the recession function of h. Their method is to approximate the nonsmooth problem by smoothing the function F using h since frequently h is smooth, but F is not. Hence this problem is approximated by the following smoothing problem

$$\inf_{x \in \mathbb{R}^n} g_\epsilon(x), \text{ where } g_\epsilon(x) = \epsilon h\left(f_1(x)/\epsilon, \ldots, f_m(x)/\epsilon\right),$$

where $\epsilon > 0$ is a parameter.

For example, if $F(y) = \sum_{i=1}^{m} |y_i|$, then it is easy to see that for $i = 1, \ldots, m$

$$\sqrt{1 + (y_i + x_i)^2} - \sqrt{1 + x_i^2} \le |y_i|$$

and

$$\lim_{|x_i| \to \infty} \left(\sqrt{1 + (y_i + x_i)^2} - \sqrt{1 + x_i^2}\right) = |y_i|.$$

Thus F is the recession function of $h(y) = \sum_{i=1}^{m} \sqrt{1 + y_i^2}$. This coincides with the smooth approximation methods given in [6, 16] and the one that we propose here. If $F(y) = \max_{1 \le i \le m} y_i$, then $h(y) = \log \sum_{i=1}^{m} y_i$. This is different from our method. We note that the approach of Ben-Tal and Teboulle depends on the recession function. In our method we assume that G is smooth and that the functions f_i satisfy some conditions, however these conditions are easily satisfied in many practical optimization problems.

We conclude this section by mentioning that in [3, 4], Bertsekas examined a two-parameter approximation for a nonsmooth minimization problem with kinks $\{0, g_i(x)\}$ and that Teo and Goh [13], Jennings and Teo [10] and Teo et al. [14] studied a class of nonsmooth optimization problems by smoothing the nondifferentiable cost functional and nondifferentiable functional inequality constraints with parameter approxima-tions.

3. Error analysis

In this section under various conditions on the function G we present error analysis and obtain estimates of the error $g(x^*) - g(x_\epsilon^*)$, where x^* is the optimal solution of (P) and x_ϵ^* is the optimal solution of (P(ϵ)). It is worth noting that error estimates are not related to the constant α.

To study error analysis, we first assume some conditions on the function G which are easily verified in many practical problems.

(A4)   The function $G(y_1, \ldots, y_m)$ is monotone in the sense that if $y_1 \geq z_1, \ldots, y_m \geq z_m$, then

$$G(y_1, \ldots, y_m) \geq G(z_1, \ldots, z_m);$$

(A5)   There exists $K > 0$ such that

$$G(y_1, \ldots, y_m) + K\epsilon \geq G(y_1 + \epsilon, \ldots, y_m + \epsilon), \qquad \forall y = (y_1, \ldots, y_m) \in \mathbb{R}^m,$$

where $\epsilon > 0$ is a smoothing parameter.

The assumption (A5) indicates that the function G increases at least as an affine function. For example, the function $G(y_1, \ldots, y_m) = \sum_{i=1}^m \beta_i y_i$ satisfies conditions (A4) and (A5) if $\beta_i \geq 0, i = 1, \ldots, m$ and if $K \geq \sum_{i=1}^m \beta_i$. It is clear that $G(x) = x^2$ is differentiable and monotone but it does not satisfy (A5). However this case will be included in Theorem 3.2.

We use the following special case of the Minkowski inequality to derive bounds on the error $g(x^*) - g(x_\epsilon^*)$.

For $a, b \geq 0, p \geq 1$, we have $(a^p + b^p)^{\frac{1}{p}} \leq a + b$.

LEMMA 3.1. *Assume that (A4) and (A5) hold. Then for every $\epsilon > 0$ and for any $x \in X$,*
$$0 \leq g_\epsilon(x) - g(x) \leq K\epsilon.$$

PROOF. Using the Minkowski inequality and noting that $f_i(x) \geq 0$ for all $x \in X$, we get

$$f_i(x) \leq (f_i(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}} \leq f_i(x) + \epsilon, \qquad \text{for} \quad i = 1, \ldots, m, \forall x \in X.$$

Thus by (A4) and (A5), we get

$$\begin{aligned}
g_\epsilon(x) &= f(x) + G\big((f_1(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}, \ldots, (f_m(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}\big) \\
&\leq f(x) + G(f_1(x) + \epsilon, \ldots, f_m(x) + \epsilon) \\
&\leq f(x) + G(f_1(x), \ldots, f_m(x)) + K\epsilon \\
&= g(x) + K\epsilon, \qquad \forall x \in X.
\end{aligned}$$

On the other hand, using (A4), we have

$$g(x) = f(x) + G(f_1(x), \ldots, f_m(x))$$
$$\leq f(x) + G\left((f_1(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}, \ldots, (f_m(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}\right)$$
$$= g_\epsilon(x), \qquad \forall x \in X.$$

Thus $0 \leq g_\epsilon(x) - g(x) \leq K\epsilon, \forall x \in X$.

We now obtain bounds for the functional values of the original objective at the approximate optimal solution and at the optimal solution under the assumptions (A4) and (A5).

THEOREM 3.1. *Assume that (A4) and (A5) hold. For any $\epsilon > 0$, we have*

$$0 \leq g(x_\epsilon^*) - g(x^*) \leq K\epsilon,$$

where x^* is a solution of (P), x_ϵ^* is a solution of (P(ϵ)).

PROOF. It follows from Lemma 3.1 that

$$g(x) \leq g_\epsilon(x), \qquad \forall x \in X.$$

Then

$$g(x_\epsilon^*) - g(x^*) \leq g_\epsilon(x_\epsilon^*) - g(x^*).$$

Since x_ϵ^* is a solution of (P(ϵ)), we have

$$g_\epsilon(x_\epsilon^*) \leq g_\epsilon(x), \qquad \forall x \in X.$$

Thus $g_\epsilon(x_\epsilon^*) \leq g_\epsilon(x^*)$. Then from Lemma 3.1 and x^* being a solution of (P) we have

$$0 \leq g(x_\epsilon^*) - g(x^*) \leq g_\epsilon(x_\epsilon^*) - g(x^*) \leq g_\epsilon(x^*) - g(x^*) \leq K\epsilon.$$

Now we derive more bounds on the error $g(x^*) - g(x_\epsilon^*)$ by relaxing the assumptions (A4) and (A5). We assume that the function G is Lipschitz. It is easy to see that a Lipschitz function may not satisfy the condition (A4). On the other hand, for a fixed parameter $\epsilon > 0$, a function satisfying (A4) and (A5) may not be Lipschitz, or even continuous.

LEMMA 3.2. *Let $G : \mathbb{R}^m \longrightarrow \mathbb{R}$ be Lipschitz with the Lipschitz constant L. Then*

$$|g_\epsilon(x) - g(x)| \leq L\sqrt{m}\epsilon, \qquad \forall x \in X.$$

PROOF. From the proof of Lemma 3.1, we obtain

$$0 \leq (f_i(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}} - f_i(x) \leq \epsilon, \quad \text{for} \quad i = 1, \ldots, m, \forall x \in X.$$

It follows from the Lipschitz condition on G that

$$
\begin{aligned}
|g_\epsilon(x) - g(x)| \\
= \left| G\left((f_1(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}, \ldots, (f_m(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}}\right) - G(f_1(x), \ldots, f_m(x)) \right| \\
\leq L \sqrt{\sum_{i=1}^{m} ((f_i(x)^\alpha + \epsilon^\alpha)^{\frac{1}{\alpha}} - f_i(x))^2} \\
\leq L \sqrt{\sum_{i=1}^{m} \epsilon^2} \leq L\sqrt{m}\epsilon, \qquad \forall x \in X
\end{aligned}
$$

Then the conclusion is proved.

THEOREM 3.2. *Let $G : \mathbb{R}^m \longrightarrow \mathbb{R}$ be Lipschitz with the Lipschitz constant L. Then*

$$0 \leq g(x_\epsilon^*) - g(x^*) \leq 2L\sqrt{m}\epsilon.$$

PROOF. It follows from Lemma 3.2 that

$$g(x) \leq g_\epsilon(x) + L\sqrt{m}\epsilon, \qquad \forall x \in X.$$

Then

$$g(x_\epsilon^*) - g(x^*) \leq g_\epsilon(x_\epsilon^*) - g(x^*) + L\sqrt{m}\epsilon.$$

Since x_ϵ^* is a solution of (P(ϵ)), we have

$$g_\epsilon(x_\epsilon^*) \leq g_\epsilon(x), \qquad \forall x \in X.$$

Thus $g_\epsilon(x_\epsilon^*) \leq g_\epsilon(x^*)$. Then from x^* being a solution of (P) and Lemma 3.2 we have

$$
\begin{aligned}
0 &\leq g(x_\epsilon^*) - g(x^*) \\
&\leq g_\epsilon(x_\epsilon^*) - g(x^*) + L\sqrt{m}\epsilon \\
&\leq g_\epsilon(x^*) - g(x^*) + L\sqrt{m}\epsilon \\
&\leq L\sqrt{m}\epsilon + L\sqrt{m}\epsilon = 2L\sqrt{m}\epsilon.
\end{aligned}
$$

The proof is complete.

4. Second-order necessary conditions for (P(ϵ))

In this section, we derive a general second-order necessary condition for the smooth approximation problem (P(ϵ)) by strengthening the assumptions (A1)–(A3). These conditions may be used to obtain second-order necessary conditions for the original problem (see [3]). We use the generalized second-order directional derivative and generalized Hessian recently introduced in [16, 17].

For simplicity, we consider the following special case ($\alpha = 2$) of the smooth approximation problem

$$\min_{x \in X} g_\epsilon(x), \text{ where } g_\epsilon(x) = f(x) + G\big(\sqrt{f_1(x)^2 + \epsilon^2}, \ldots, \sqrt{f_m(x)^2 + \epsilon^2}\big),$$

where $\epsilon > 0$ is a smoothing parameter. We still call it (P(ϵ)). The results obtained in this section hold for the case $\alpha > 1$ with modified versions. To obtain second-order necessary conditions of (P(ϵ)), we strengthen the conditions (A1)–(A3) by assuming:

(A1)′ $f : X \longrightarrow \mathbb{R}$ is $C^{1,1}$;

(A2)′ $G : \mathbb{R}^m \longrightarrow \mathbb{R}$ is $C^{1,1}$;

(A3)′ $f_i : X \longrightarrow \mathbb{R}, i = 1, \ldots, m$ are not, in general, $C^{1,1}$, but the square of the functions f_i, f_i^2, $i = 1, \ldots, m$, are $C^{1,1}$.

It is well-known that the square of the maximum of two twice differentiable functions may not be twice differentiable, but is $C^{1,1}$ under some conditions. Let $F(x) = (F_1(x), \ldots, F_m(x)) = (f_1(x)^2, \ldots, f_m(x)^2) : X \longrightarrow \mathbb{R}^m$ and

$$h(y_1, \ldots, y_m) = G\big(\sqrt{y_1 + \epsilon^2}, \ldots, \sqrt{y_m + \epsilon^2}\big) : O \subset R^m \longrightarrow \mathbb{R},$$

where O is an open subset of $\mathbb{R}^m$. Then $g_\epsilon(x) = f(x) + (h \circ F)(x)$.

THEOREM 4.1. *Assume that (A1)′–(A3)′ hold. If z is a local minimum of (P(ϵ)), then for any $u \in X$*

$$f^{\infty}(z; u, u) + h^{\infty}(F(z); \nabla F(z)u, \nabla F(z)u)$$

$$+ \frac{1}{2} \max \left\{ \sum_{i=1}^{m} \frac{\partial h(F(z))}{\partial x_i} \frac{1}{\sqrt{f_i(x)^2 + \epsilon^2}} \langle y_i^*, u \rangle : y_i^* \in \partial^{\infty} F_i(z)(u) \right\} \geq 0.$$

PROOF. It follows from a second-order necessary condition (see Section 1) that

$$g_\epsilon^{\infty}(z; u, u) \geq 0, \qquad \forall u \in X.$$

Thus

$$f^{\infty}(z; u, u) + (h \circ F)^{\infty}(z; u, u) \geq 0, \qquad \forall u \in X.$$

We now calculate $(h \circ F)^{\infty}(z; u, u)$ using the generalized chain rule for composite $C^{1,1}$ functions (see Section 1). For any $u \in X$,

$$(h \circ F)^{\infty}(z; u, u)$$
$$\leq h^{\infty}(F(z); \nabla F(z)u, \nabla F(z)u)$$
$$+ \max\left\{ \sum_{i=1}^{m} \frac{\partial h(F(z))}{\partial y_i} \langle y_i^*, u \rangle : y_i^* \in \partial^{\infty} F_i(z)(u) \right\}$$
$$= h^{\infty}(F(z); \nabla F(z)u, \nabla F(z)u)$$
$$+ \max\left\{ \sum_{i=1}^{m} \frac{\partial h(F(z))}{\partial x_i} \frac{\partial x_i}{\partial y_i}\bigg|_{y_i = f_i(z)^2} \langle y_i^*, u \rangle : y_i^* \in \partial^{\infty} F_i(z)(u) \right\}$$
$$(x_i = \sqrt{y_i + \epsilon^2})$$
$$= h^{\infty}(F(z); \nabla F(z)u, \nabla F(z)u)$$
$$+ \max\left\{ \sum_{i=1}^{m} \frac{\partial h(F(z))}{\partial x_i} \frac{1}{2} \frac{1}{\sqrt{f_i(z)^2 + \epsilon^2}} \langle y_i^*, u \rangle : y_i^* \in \partial^{\infty} F_i(z)(u) \right\}.$$

Then the conclusion holds.

COROLLARY 4.1. *Assume that* (A1)′–(A3)′ *hold and* $G(y_1, \ldots, y_m) = M \sum_{i=1}^{m} y_i$. *If* z *is a local minimum of* $(P(\epsilon))$, *then for any* $u \in X$

$$f^{\infty}(z; u, u) + \frac{M}{2} \sum_{i=1}^{m} \frac{1}{\sqrt{f_i(z)^2 + \epsilon^2}} F_i^{\infty}(z; u, u) \geq 0.$$

PROOF. Now we have

$$h^{\infty}(F(z); \nabla F(z)u, \nabla F(z)u) = 0, \qquad \frac{\partial h(F(z))}{\partial x_i} = M, \qquad i = 1, \ldots, m.$$

Then the conclusion follows from Theorem 4.1.

COROLLARY 4.2. *If in addition to the assumptions in Corollary 4.1, the functions* f *and* F_i *are twice weakly Gâteaux differentiable at* z, *then for any* $u \in X$

$$\langle \nabla^2 f(z)u, u \rangle + \frac{M}{2} \sum_{i=1}^{m} \frac{1}{\sqrt{f_i(z)^2 + \epsilon^2}} \langle \nabla^2 F_i(z)u, u \rangle \geq 0.$$

PROOF. Since f is twice weakly Gâteaux differentiable, we get (see [17])

$$f^{\infty}(z; u, u) = \langle \nabla^2 f(z)u, u \rangle.$$

Hence the result is proved by using Corollary 4.1.

Note that $F_i^{\infty\infty}(z; u, u)$ may not be easily calculated for general functions f_i. In the following, we give some examples where $F_i^{\infty\infty}(z; u, u)$ can be easily calculated. These functions are $d_C(x)$ and $\max\{g_1(x), 0\}$.

1.   Let $f_1(x) = d_C(x)$, where C is a closed convex set of a Hilbert space X. Then $F_1(x) = f_1(x)^2$ is $C^{1,1}$ (see Holmes [8]). It was shown in [16] that if $f_1(x)$ is twice semi-regular at $z \in C$, then

$$F_1^{\infty\infty}(z; u, u) = 0, \qquad \forall u \in \text{cone}(C - z).$$

2.   Let $f_1(x) = \max\{g_1(x), 0\}$, where $g_1(x)$ is a $C^{1,1}$ function on X. Then it is easy to see that $F_1(x) = f_1(x)^2$ is $C^{1,1}$. It follows from the chain rule (see Section 1) that for every $u \in X$

$$F_1^{\infty\infty}(z; u, u) = \begin{cases} 2g_1(x)g_1^{\infty\infty}(z; u, u) + 2\langle \nabla g_1(z), u \rangle^2, & \text{if } g_1(z) \geq 0, \\ 0, & \text{if } g_1(z) < 0. \end{cases}$$

This result was given in [16] when $g_1(x)$ is twice differentiable.

5. Conclusion

We have shown that certain types of composite nonsmooth optimization problems can be approximated by a smooth optimization problem and have provided a unified study of the smooth techniques presented recently in [2, 6, 16]. We presented error analysis of the function value at optimal solution and at approximation optimal solution under various conditions. We also derived second-order necessary conditions for smooth approximation problems using a recently introduced generalized second-order directional derivative. These conditions may be useful to study second-order necessary conditions of the original minimization problems.

We conclude by giving an example to show the usefulness of the results. Consider the minimization problem:

$$\min_{x \in \mathbb{R}} f_1(x),$$

where $f_1(x) = |x|^{1/3}$. It is clear that $x^* = 0$ is an optimal solution and the problem is not smooth. This problem can be approximated by the following smooth problem

$$\min_{x \in \mathbb{R}} (f_1(x)^6 + \epsilon^6)^{1/6} = (x^2 + \epsilon^6)^{1/6},$$

where $\epsilon > 0$ and $\alpha = 6$. It is easy to see that the approximation problem has optimal solution $x_\epsilon^* = 0$. Thus the error estimates given in Theorem 3.1 and Theorem 3.2 are trivially satisfied and second-order necessary condition in Theorem 4.1 also holds.

Acknowledgements

References

[1] A. Ben-Tal and M. Teboulle, "A smoothing technique for nondifferentiable optimization problems", in *Optimization - Fifth French-German Conference* (ed. S. Dolecki), Volume 1405 of *Lecture Notes in Math.*, (Springer-Verlag, New York, 1989) 1–13.

[2] A. Ben-Tal, M. Teboulle and W. H. Yang, "A least-square method for a class of nonsmooth minimization problems with applications in plasticity", *Applied. Math. Optimiz.* 24 (1991) 272–288.

[3] D. P. Bertsekas, "Nondifferentiable optimization via approximation", *Math. Prog. Study* 3 (1975) 1–25.

[4] D. P. Bertsekas, *Constrained optimization and lagrange multiplier methods* (Academic Press, New York, London, 1982).

[5] F. H. Clarke, *Optimization and nonsmooth analysis* (John Wiley, New York, 1983).

[6] R. A. El-Attar, M. Vidyasagar and R. K. Dutta, "An algorithm for l_1-norm minimization with application to nonlinear l_1 approximation", *SIAM J. Numer. Anal.* 16 (1979) 70–86.

[7] R. Fletcher, *Practical methods of optimization* (John Wiley, New York, 1987).

[8] R. B. Holmes, *A course on optimization and best approximation* (Springer-Verlag, New York, 1972).

[9] A. D. Ioffe, "Necessary and sufficient conditions for a local minimum, 2: conditions of Levitin-Milijutin-Osmoloviskii type", *SIAM J. Control Optimiz.* 17 (1979) 245–250.

[10] L. S. Jennings and K. L. Teo, "A computational algorithm for functional inequality constrained optimization problems", *Automatica* 26 (1990) 371–375.

[11] V. Jeyakumar, "Composite nonsmooth programming with Gâteaux differentiability", *SIAM J. Optimiz.* 1 (1991) 30–41.

[12] R. A. Polyak, "Smooth optimization methods for minimax problems", *SIAM J. Control Optimiz.* 26 (1988) 1274–1286.

[13] K. L. Teo and C. J. Goh, "On constrained optimization problems with nonsmooth cost functionals", *Applied Math. Optimiz.* 18 (1988) 181–190.

[14] K. L. Teo, V. Rehbock and L. S. Jennings, "A new computational algorithm for functional inequality constrained optimization problems", *Automatica* 29 (1993) 789–792.

[15] X. Q. Yang, "Second-order conditions in $C^{1,1}$ optimization with applications", *Numer. Funct. Anal. Appl.* 14 (5&6) (1993) 621–632.

[16] X. Q. Yang, "Generalized second-order directional derivatives and optimality conditions", *Nonlinear Analysis — TMA* (to appear) .

[17] X. Q. Yang and V. Jeyakumar, "Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions", *Optimization* 26 (1992) 165–185.